scientific reports

Check for updates

OPEN

Oscillating spatiotemporal patterns of COVID-19 in the United States

Hawre Jalal[1]✉, Kyueun Lee[2] & Donald S. Burke[3]

COVID-19 case rates in the US wax and wane in wave-like patterns over time, but the spatial patterns of these temporal epidemic waves have not been well characterized. By analyzing state- and county-level COVID-19 case rate data for spatiotemporal decomposition modes and oscillatory patterns, we demonstrate that the transmission dynamics of COVID-19 feature recurrent spatiotemporal patterns. In addition to the well-recognized national-level annual mid-winter surges, we demonstrate a prominent but previously unrecognized six-month north–south oscillation in the eastern US (Eastern US COVID-19 Oscillator—EUCO) that gives rise to regional sub-epidemics and travelling epidemic waves. We also demonstrate a second less prominent pattern that oscillates east–west in the northern US (Northern US COVID-19 Oscillator—NUCO). The drivers of these newly recognized oscillatory epidemic patterns remain to be elucidated.

Since the first documented case of SARS-CoV-2 in the USA in January of 2020, the virus has caused nearly 100 million reported infections and more than 1 million deaths[1]. COVID-19 incidence rates have waxed and waned, with the dominant nation-wide pattern showing increases in the winter months. In mapping case incidence rates, we noticed month-to-month variations in intensity across different regions of the country. Furthermore, maps of case incidences in the second year of the epidemic appeared to show similar spatiotemporal patterns to those in the first year[2]. Intrigued by these preliminary visual observations, we set out to use computational methods to analyze the data for patterns.

The spatiotemporal granularity of available COVID-19 data in the USA provides an unprecedented opportunity to investigate detailed epidemiological dynamics of an infectious disease. In this study, we examined this spatiotemporal disease incidence data to reveal heretofore unrecognized sub-national oscillatory disease patterns. At this point in time, we do not have a sound mechanistic explanation for these newly discovered oscillations. Characterization of these repeating patterns may aid in understanding of infectious disease dynamics, enable forecasting of future surges, and facilitate improved targeting of public health resources.

Results

To better understand the transmission dynamics of COVID-19 in the USA, we carried out several distinct but complementary analyses, including hierarchical clustering, cross-date rank correlations, and singular value decomposition into component modes using COVID-19 case and death data from the New York Times. For our analysis we used data from 937 days of the pandemic until the end of our analysis period on August 15, 2022. Although the New York Times provides data beyond this date through early 2023, the accuracy and frequency of reporting declined during the latter part of this period. For more details on the approaches, please refer to the Methods section below.

North–south oscillations of COVID-19 using state-level data

Figure 1 provides evidence for a prominent north–south oscillating pattern of COVID-19 case rates in the USA. Figure 1A displays a hierarchical clustering dendrogram of state-by-state similarities of case rate time-series patterns for the 48 contiguous continental states. Geo-contiguous and nearby states show very similar patterns. The dendrogram also characterizes states as belonging to the main northern (purple) or the main southern (green) regional clusters, as mapped in Fig. 1B. A boundary between the north and south case rate pattern clusters lies roughly along 37°–38° north latitude. To extend this analysis of spatial patterns, we analyzed temporal patterns by computing a matrix that compares the rank order of the 48 continental states' COVID-19 case rates on any given date on the x-axis to the rank order of state COVID-19 case rates on any other date on the y-axis (Fig. 1C). We refer to this matrix as cross-date state rank correlation matrix. This cross-date correlation

[1]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON K1G 5Z3, Canada. [2]The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, University of Washington, Seattle, WA 98195, USA. [3]Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA 15261, USA. ✉email: hjalal@uOttawa.ca

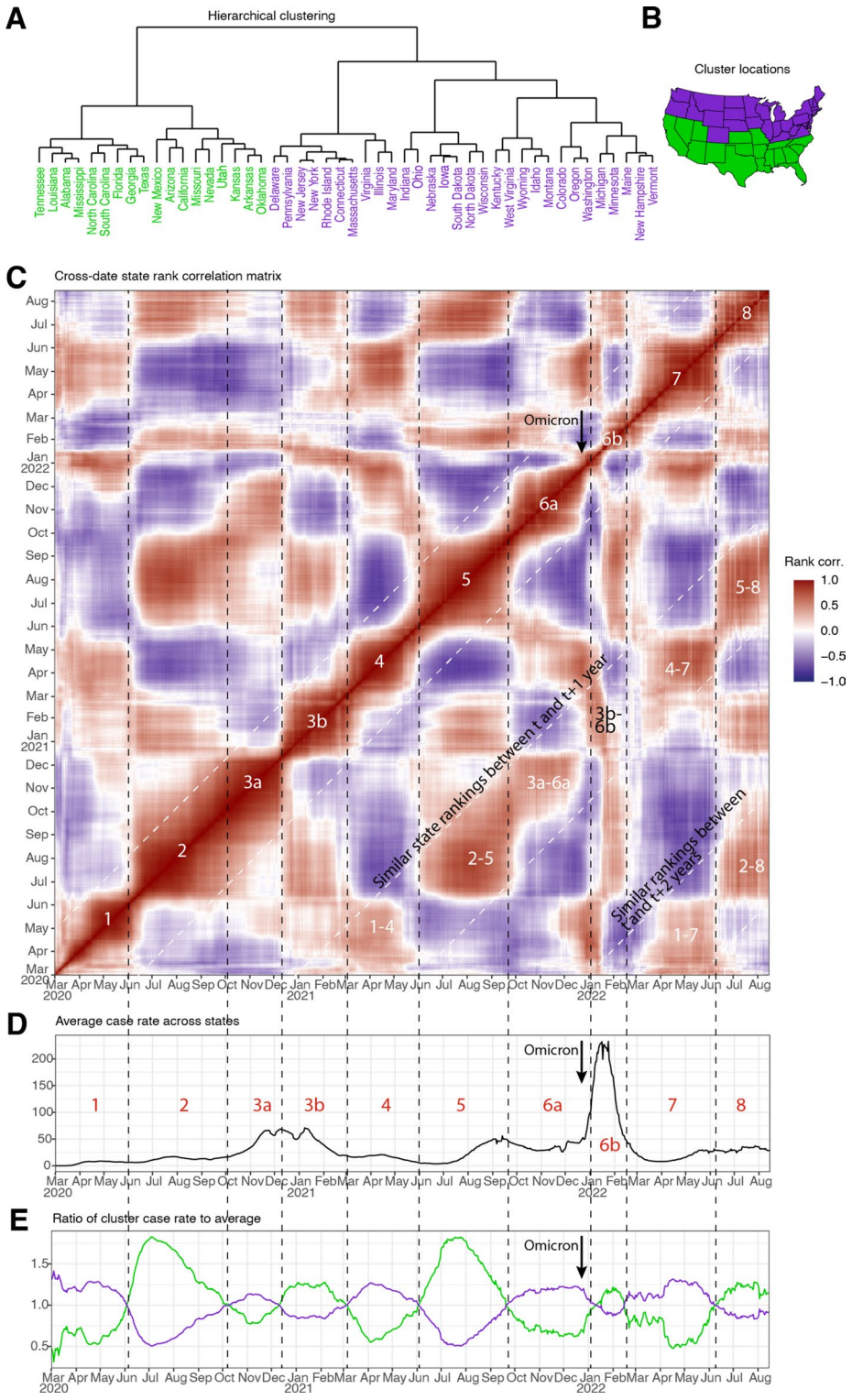

Fig. 1. North-south oscillatory pattern from state-level incidence data. **(A)** Hierarchical clustering of states using log case rate time series over the 937 day observation period shows a northern and a southern cluster. **(B)** North-south separation of states within each of the two major clusters displayed on a map. **(C)** Matrix of "cross-date" state rank correlations showing similar ranking of states within roughly 3-month periods that predictably recur. Each pixel in the 937 days x 937 days matrix represents the rank correlation coefficient of case rates across states between a date on the x-axis another date on the y-axis. Ten time-blocks of high rank correlations appear on the diagonal. These blocks represent waves 1, 2, 3a, 3b, 4, 5, 6a, 6b, 7 and 8. In addition, to high-correlation blocks on the diagonal representing close dates, two bands parallel to the diagonal at t+1 and t+2 years are marked to display the similarity in the block-patterns at the next 1 and 2 years respectively. For example, "2–5" indicates similarly in state-ranking between waves 2 and 5. **(D)** Shows the aggregated pattern of the average case rate across all 48 states. **(E)** Displays the ratios of case rates in the northern (purple) and southern (green) clusters relative to the overall average in **(D)**. The vertical lines traversing **(C– E)** indicate the approximate dates on which case rates between the northern and southern clusters cross. These lines also correspond to the boundaries of the diagonal blocks with homogeneous high cross-date correlations in **(C)**. The arrow represents the start of Omicron wave in the USA.

matrix generates a structured "checkerboard" pattern. As expected, high correlations appear along the main matrix diagonal, indicating that any two or more consecutive days typically share very similar rank orderings of state case rates. However, there are unexpected results: First, this matrix displays a steady repeating pattern of block-like periods of internally high cross-date state rank correlations, all along the diagonal, with each homogeneous block pattern lasting about three months. We introduce a terminology of numbering each of these highly spatiotemporal internally correlated blocks from earliest (1) to most recent (8). National temporal wave 3 can be seen in the cross-date state rank correlation matrix as being composed of two distinct blocks, which we label 3a and 3b to retain reference to the national level temporal waves. Similarly, national temporal waves 6 can be seen to be composed of two blocks, 6a and 6b.

Second, there are at least two more diagonal bands of high correlations parallel to the main diagonal, the first at one year later (t + 1), and a second now forming at two years later (t + 2). These parallel diagonal bands result from similar state case rate rankings at almost exactly one year and two years earlier, respectively. Third, there are additional correlated sub-patterns at t + 0.5 and t + 1.5 years. For example, states have similar rankings at the intersection of waves 2 and 3b; both of which had high case incidences in the south. We observed that overall orderly repeating "checkerboard" pattern was transiently disrupted by the introduction of the Omicron variant of SARS-CoV-2, when all regions of the country were affected simultaneously and severely (see below). After Omicron, the regular checkerboard epidemic pattern then spontaneously restored itself, with waves 7 and 8 showing the expected state rank cross-date correlation time blocks (wave 7 correlating with waves 1 and 4, and wave 8 correlating with waves 2 and 5, respectively). Finally, we found that the relative intensity of COVID-19 cases in the northern US versus the southern US (compared to the national average) oscillates every three months, aligning well with the durations of cross-state correlation blocks in the checkerboard (Fig. 1E). This indicates that the regular block patterns in the cross-date state rank correlation matrix represent the oscillating north/south changes in case rates by latitude. Accordingly, we used the dates of these normalized incidence crossing points to define wave boundaries.

We confirmed this clear recurring pattern of case rates by creating a similar matrix of cross-date rank correlations using death rates instead of case rates (Fig. 2). The COVID-19 state death rate cross-date rank order correlation matrix generates a checkerboard pattern that is almost identical to the state case rate pattern, though not as sharp due to the lower rates of mortality compared to incidence, and with a lag of about one month compared to the timing of the case rate oscillations.

Oscillations detected by decomposition of county time series data

Given the evidence for north–south oscillations in US COVID-19 case rates using state-level data, we next sought to refine the spatiotemporal patterns of these oscillations by analyzing more granular county-level case rate time-series data. We used singular value decomposition (SVD)—a method which has extensive use in signal processing and machine learning—to decompose county-level daily case rates into lower rank fundamental components (i.e. modes).

Results of the SVD are shown in Fig. 3. The first four SVD decomposition modes accounted for more than 85% of the total US log-transformed case rate variances by county and over time (Fig. 3A). We chose to analyze just these four modes in greater detail because after Mode IV the proportion of explained variance declined abruptly to 1%.

Figure 3B shows the four modes' temporal patterns (eigen-trends). These patterns represent the trends of COVID-19 case rates across all counties for each day. The contribution of these trends in each county vary by a set of weights. Figure 3C shows the modes' spatial patterns (eigen-maps) that reflect the correlation of each of the temporal patterns with each county's observed case rate over time. Red and blue areas represent counties where the mode is strongest and white areas represent counties where the correlation is weak. For example, Mode II is strongest in the northeast (high negative correlation) and the southeast (high positive correlation) compared to the counties along the central latitude.

Modes I through IV each defines a distinct spatiotemporal pattern. Mode I is the strongest (explaining 75% of the total variance) as expected because this mode captures the variance of the time-trend in case rates across all counties. The temporal trend of this mode (Fig. 3B) follows a generally cyclical annual repeating pattern with a fall-winter spike with other variable rises and falls. Modes II, III and IV capture additional local variations in the time-trend.

Figure 3D is a panel of time series heatmaps that show all county-level COVID-19 case rate time series trajectories according to day (x axis), county latitude (y axis), county longitude (five columns), and mode (five rows). The first row shows the observed data, and the following rows represent the predicted case rates for each county produced by multiplying the temporal-trends (Fig. 3B) by the county specific weights underlying the spatial weights shown in Fig. 3C.

Mode I in Fig. 3D shows a vertical band-like pattern that reflects the same overall temporal wave-pattern observed in Mode I in Fig. 3B. However, the intensity of these waves vary by county because each county has its own unique weight-adjusted magnitude for this mode's temporal trend. As expected, cases across all counties follow the average temporal trend.

Unlike Mode I, Mode II reveals a clear north–south oscillation of COVID-19 case rates in the eastern US (East of 100 degree W). We refer to this Mode as the Eastern US COVID-19 Oscillation; EUCO. This mode accounts for 5.2% of the total variance (Fig. 3A). In the spatial weight, this mode correlates most positively with the southeast region and negatively with the northeast region (Fig. 3C). As a result, Mode II captures a strong north/south oscillatory behavior, most prominent in the three longitudinal bands east of latitude 100W (Figs. 3D and 4A). This oscillatory behavior alternates between a high case rate in waves 1, 3a, 4, 6a and 7 in the northeast and waves 2, 3b, 5, 6b and 8 in southeast. This north–south oscillating pattern effectively locally tilts

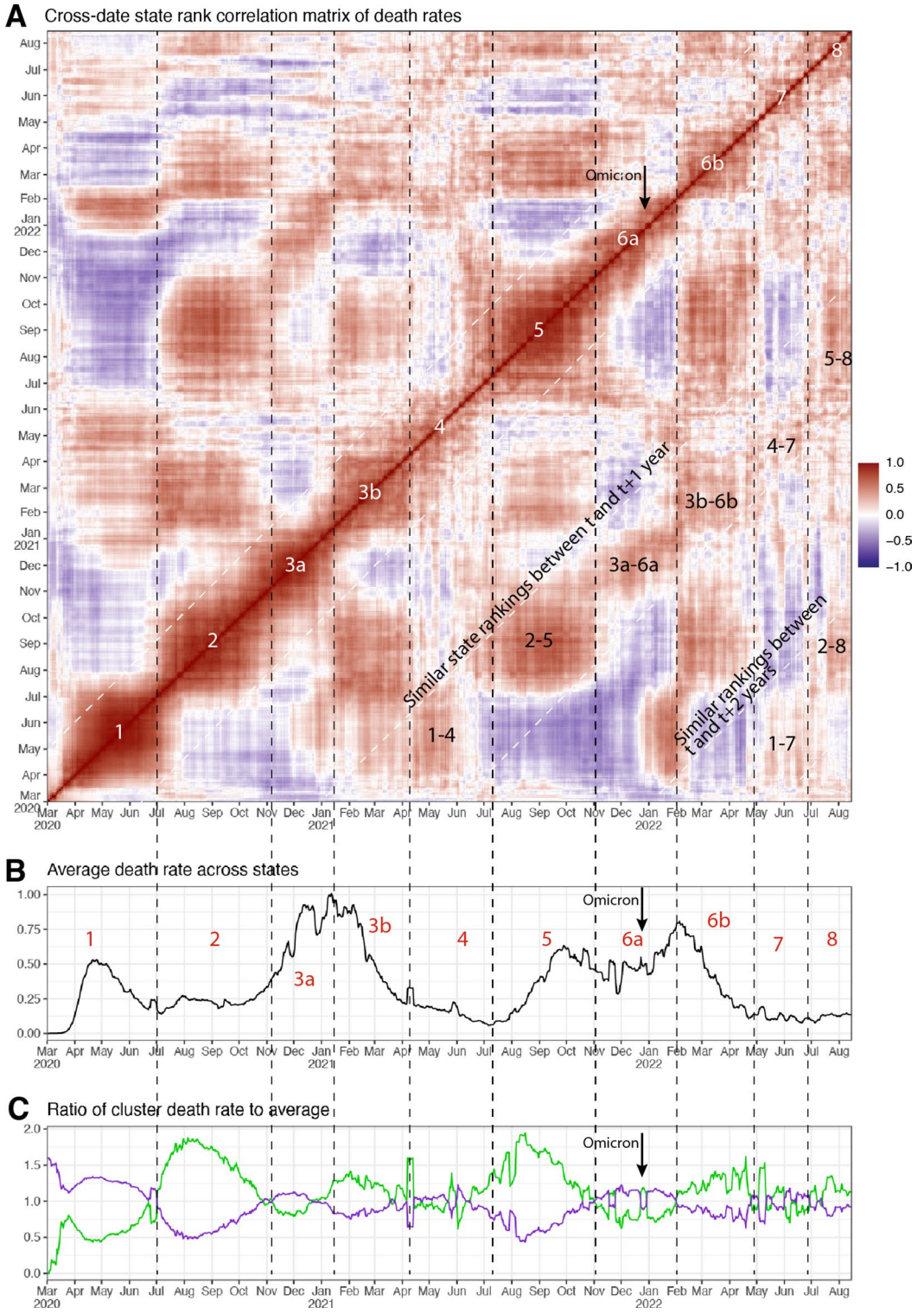

Fig. 2. State-level analysis of COVID-19 mortality confirms the north–south oscillation observed using case data. Panel A is a matrix of cross-date state rank correlations showing similar ranking of state mortality rates within each 3-month periods. Panel B shows the average mortality rate aggregated across states, and Panel C shows the average mortality rate within the northern and southern state clusters defined in Fig. 1A. The patterns in (**A–C**) are overall similar to those for case rates shown in Fig. 1, panels (**C–E**) respectively, but the mortality patterns are lagged by about one month and are less well defined especially toward the later time periods due to declining case fatality ratio (notice the relatively low ratio of Omicron mortality to cases).

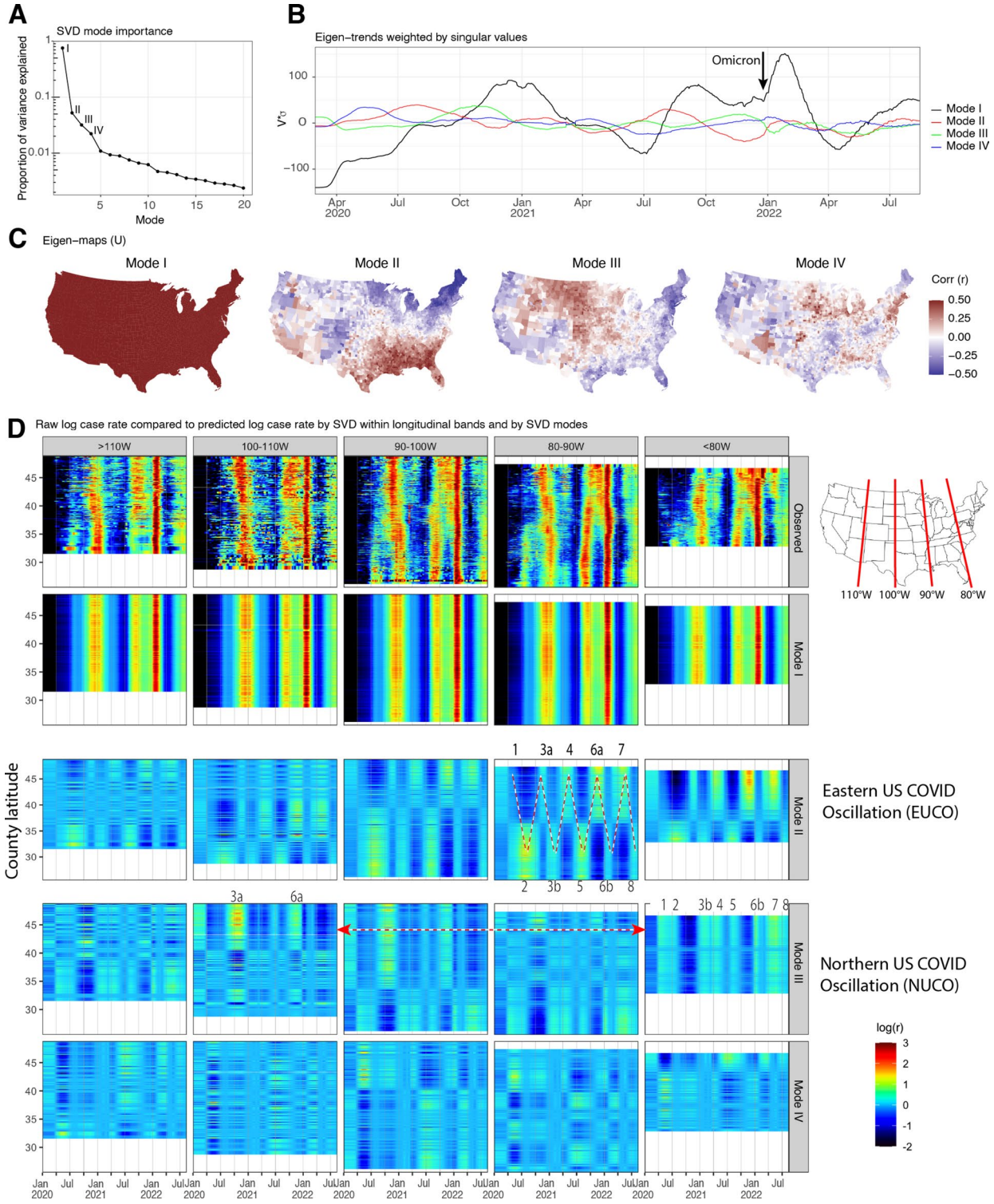

Fig. 3. Singular value decomposition (SVD) of county-level data confirms the state-level north–south oscillatory patterns and defines its regional extent. (**A**) Shows the proportion of the variance explained by the top 20 modes. The first four modes are labelled with Roman Numerals I, II, III and IV. (**B**) Shows the temporal patterns of the modes by plotting the eigen-trends of the SVD modes weighted by their singular values. The start of Omicron (arrow) was associated with a sudden increase in cases in Mode I. (**C**) Shows the spatial patterns of the modes by showing the correlation of the eigen-maps of the SVD modes with each county's log case rate. (**D**) Compares the observed log case rates (top row) against the predicted case rate from each mode by latitude within five longitudinal bands as indicated in the legend on the right. The Eastern US COVID-19 Oscillation (EUCO) can be seen across the USA but is most prominent in the eastern longitudinal bands. The north–south alternations of the EUCO are shown with dashed lines in the Mode II 80-90W panel. In addition, the Northern US COVID-19 Oscillation (NUCO) is illustrated for Mode III alternating between the northeast (longitudes < 80) and north-central (longitudes 90–100) labelled with a horizontal arrow.

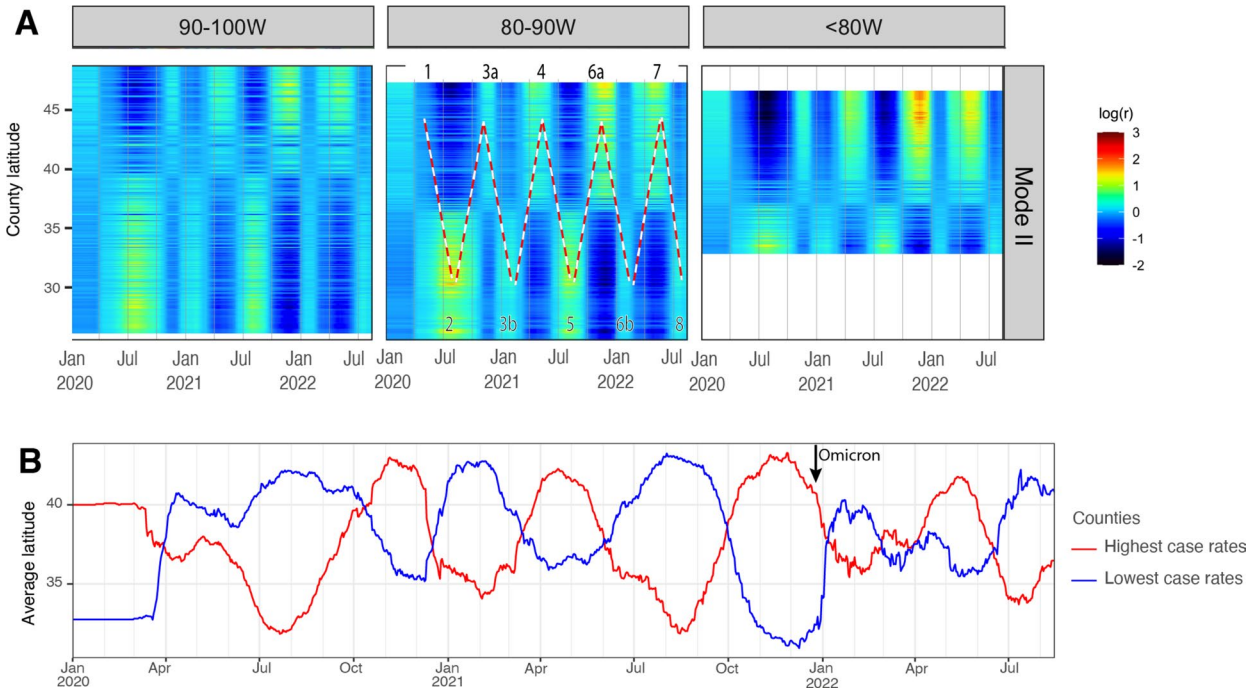

Fig. 4. The eastern US COVID-19 oscillation (EUCO). (**A**) Zooms in on the SVD Mode II for the eastern US (east of 100 W longitude) from Fig. 3. The EUCO shows as an alternating pattern of north–south intensity back and forth across 37 degrees north latitude. In the upper center panel red dashed lines are used to show this alternation in case rates for counties in the 80–90 W longitude band. The same pattern can be seen for counties in the < 80 W and the 90–100 W longitudinal bands. (**B**) Reveals the EUCO in the raw data by showing the oscillation of the average latitude among counties with the highest and lowest daily case rates in the eastern US.

the time-invariant patterns of Mode I up and down along the axis of 37 degrees north latitude so that the waves behave as back-and-forth travelling waves across latitude.

The EUCO is further documented in Fig. 4B by plotting the daily average latitude among the counties with the highest versus the lowest case rates, among all counties east of longitude 100 W. This figure reveals large semi-annual reciprocally alternating north–south latitudinal swings of the locations of the counties with the lowest and the highest case rates. Although these large oscillations were partially disrupted by the Omicron wave in winter 2022, they appear to have bounced back to their original pattern since April 2022.

Similarly, Mode III reveals a northern US COVID-19 oscillation (NUCO). This mode explains 3.2% of the variance of the overall case rates and correlates positively with the fall waves in the northcentral US counties, and negatively correlates with winter, spring and summer waves (Fig. 3B, C). This mode captures high case rates in the fall (waves 3a and 6a) in longitudinal bands 90–100 W and 100–110 W, and to a lesser degree, high cases in the winter, spring and summer in the northeast—waves 1, 2, 3b, 4, 5, 6b, 7 and 8 in longitudinal band < 80 W (Fig. 3D).

Lastly, Mode IV explains 2.2% of the variance. This mode is correlated the most with the spring wave in the northeast, and to a lesser extent with case rates in the upper Midwest and southeast (Fig. 3C, D). Overall, this mode reflects the change in case rate severity in the first annual cycle relative to the second and third cycles (Fig. 3D).

Although Modes II–IV explain 11% of the variance compared to 75% for Mode I, most of Mode I's captured variance occurs during the large epidemic winter peaks compared to Modes II–IV that have distinct contributions to waves in the other times of the year.

Consistent with Fig. 2, Mode II shows an alternating pattern of regional high intensity in the north and the south. When these high intensity Mode II regional patterns occur during the nadirs of Mode I, the Mode II localized epidemic is the dominant contributor to regional county case rates. This has occurred in waves 2, 4, 5, and 7. Mode III suggests that there may be an annual northeast-northcentral oscillation where case rates are highest in the northcentral region in the fall waves (3a and 6a) and highest in the northeast in the spring waves (4 and 7). Mode III was especially prominent in wave 3a. It is difficult to discern any repeating patterns of Mode IV, but this mode did have the highest intensities in waves 1 and 5.

We focused our analyses presented here on spatiotemporal patterns by latitude to generate four SVD modes. These four modes added together closely reproduce the observed case rate patterns in all five of the US regions analyzed (Fig. 5).

Note that these oscillatory patterns in COVID peak rates generate travelling epidemic waves. In Fig. 5, as shown for example for the observed cases in the longitudinal band B (80—90 W), the sequential waves 3a and 3b move gradually from north to south over time, while wave 5 travels from the south to the north. Wave 6a begins

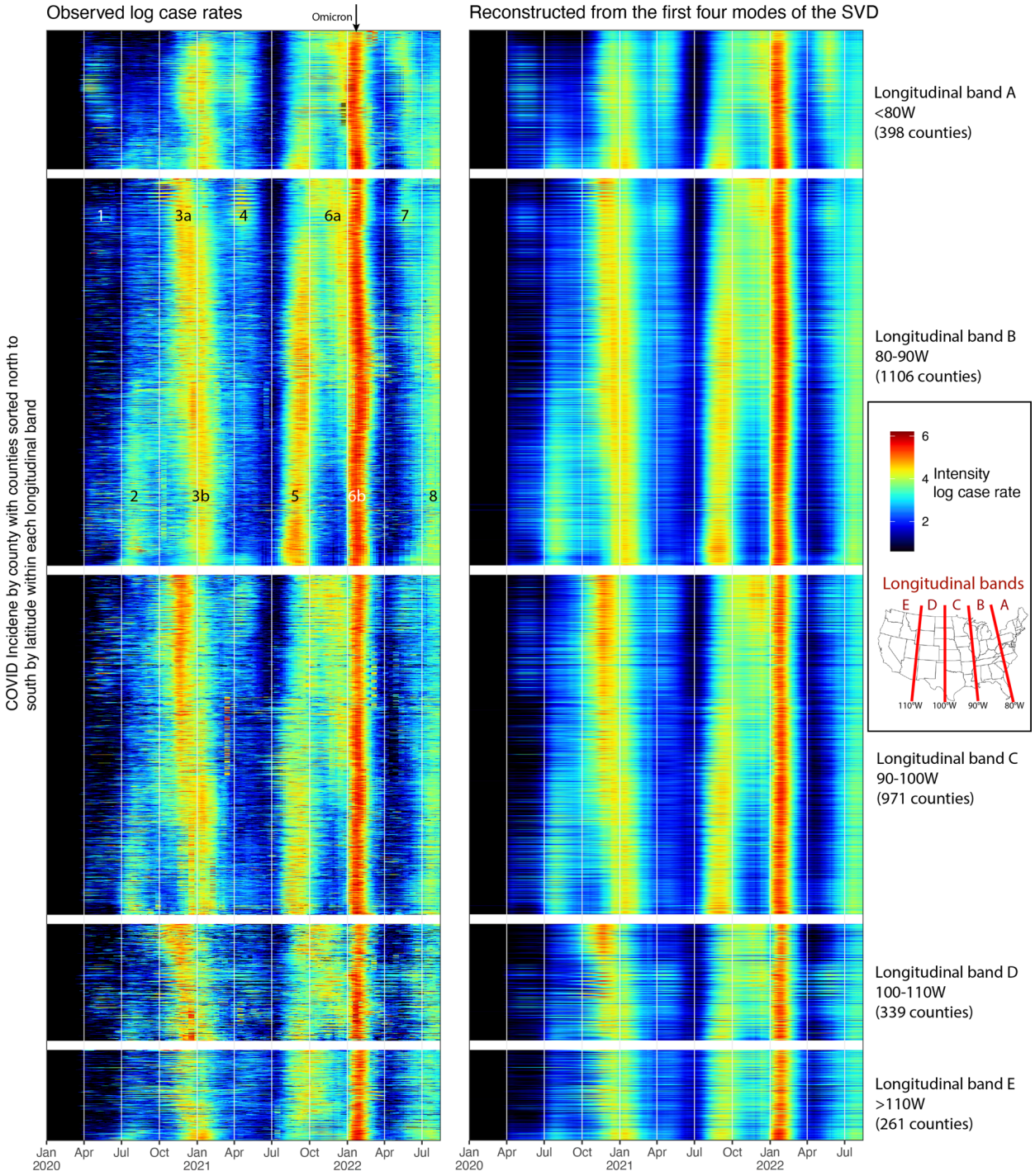

Fig. 5. The first four singular value decomposition (SVD) modes together closely reconstruct observed case rates by latitude within longitudinal bands. The panels in the left column represent heatmaps of real observed log case rates [log(r)]. The x-axis represents time, and the y-axis represents all US counties ranked by their latitude from northernmost to the southernmost within each of the five longitudinal bands. Temporal waves 1–8 are labelled on the panel of observed log case rates for longitudinal band 80–90 W. A slanted left/right or right/left pattern of a wave shows it to be a travelling wave. The right column of panels shows the reconstruction of log case rates from the first four modes (Modes I, II, III and IV) of the singular value decomposition (SVD) analysis for all five longitudinal bands. Notice the Omicron wave in early 2022 (arrow) partially overlapping wave 6a.

in the north and moves southward, but the near-simultaneous nationwide emergence of Omicron obscures the expected 6b component of the travelling wave.

In addition to these patterns across latitudes, we also found limited but less-striking evidence of additional spatial structure by longitude within latitudinal bands.

Discussion

Case reporting during COVID-19 epidemic has generated the most massive and detailed spatiotemporal data set in the history of epidemiology. In the USA, approximately 3 million case rate data points have been recorded in detailed time and space (937 daily case counts from 3108 counties), tallying the date and location of a cumulative 95 million cases[1]. Although reporting can be incomplete and imperfect, we were able to mine this rich spatiotemporal data similar to other studies that have employed advanced computational methods[3–6]. By detecting and describing heretofore unrecognized recurring patterns we provide new evidence that the COVID-19 epidemic in the USA has recurring spatiotemporal components.

Our most striking finding was the alternating north–south intensity of COVID-19 case rates in the Eastern US [Eastern US COVID-19 Oscillation (EUCO)]. The SVD Mode II captures the variance generated by this heretofore unrecognized north–south oscillation, with case rates oscillating around a mid-point of approximately 37-degree north latitude. Although mode II accounts for only 5.2% of the total national epidemic-long variance, this percentage can be considered to be minimized because in the SVD the case rates are log-transformed rather than the natural scale of case rates. Nonetheless, mode II accounted for the majority of the regional variance in the northeast and southeast regions during waves 2, 4, 5, 6a, and 7, that is during regional waves that were not dominated by the larger nation-wide mode I-associated winter waves (Fig. 3).

While the timing of COVID-19 winter waves is consistent with that of other respiratory viruses, the timing and location of the summer/southern waves 2, 5 and 7 is unexpected. These waves start near the southern US border, the hottest time of the year when relative humidity is high. Factors associated with southern/summer increased transmission might include increased indoor gatherings during hot seasons, exposure to cool dry air from air-conditioners, or other unknown factors[7,8]. Further analyses are warranted.

Similarly, mode III appears to follow an oscillatory pattern [Northern US COVID-19 Oscillation (NUCO)], with waves in the northeast (< 80° W) displaying high rates for most of the year, which are transiently interrupted by a sharp drop in rates during the fall wave 3a and another sharp nadir one year later in wave 6a. This temporal pattern of case rates is exactly the opposite that of the north central region (90°–110° W) which experienced transient sharp surges of higher rates in waves 3a and 6a.

These recurring patterns show that COVID-19 waves can have sub-national travelling wave patterns, similar to the patterns seen for some other infectious diseases[9–12]. These pattern analyses also suggest that the oscillating north/south wave patterns we report here may not be confined to the USA, but may be part of a larger North American continental pattern. This US-based analysis should be extended to the entire North American continent.

In this study, we focused exclusively on defining the patterns of COVID-19 itself and did not attempt to identify the specific mechanistic drivers of these patterns. Many factors are known or thought to be associated with waxing and waning in COVID-19 case rates, including the introduction of new virus variants [13]; the level of population immunity, either through infection or vaccines [14]; human mobility and public health interventions such as social distancing and masks [15,16]; and environmental factors such as ambient or indoor temperature and humidity [17–22]; and other factors [23,24]. Thus, epidemic forecasting is challenging because all of these factors are continuously at play in epidemic dynamics, and only occasionally does one factor rise in importance to dominate the picture[25,26].

Annual cycling of respiratory virus transmission, including common cold coronaviruses, is well documented[27,28], so the rough annual frequency traced by the two national COVID case rate curves (SVD mode I) was not surprising. However, our finding of the additional EUCO and NUCO regional oscillators of COVID-19 case rates was unexpected. Additionally, our observation that COVID-19 case rates repeatedly increased each summer in the southern US was unexpected, as this observation runs counter to the expected associations of SARS-CoV-2 transmission and colder weather. If the south had a single peak each year, it might indeed be possible to explain the oscillation by the annual temperature cycle, but the south had not just one but two annual peaks, one in the hottest months of July and August, and the other in January and February. The surge of cases in the south in the summer is contrary to the expected seasonal surges of common respiratory viruses, and inconsistent with the hypothesis that they are related to colder temperatures in the conventional way. Isolation of mode II data as we have done here, and essentially discarding other mode data as noise, should facilitate studies to find the mechanistic drivers of this strong north–south epidemic oscillator.

Our finding that there are regional oscillatory rhythms in the COVID-19 epidemic may have implications for evaluation of control policies. Any given rise or fall in case rates in a jurisdiction may be associated with the endogenous oscillations we describe here, rather than with a newly implemented policies in that jurisdiction. COVID-19 policy analyses should be appropriately controlled to take into consideration the expected EUCO and NUCO rises and falls. At this point, we do not have a mechanistic explanation for either oscillation. The oscillations can be driven by an increase in the rate of new infections, or by an increase in severity of the cases leading to an apparent increase in case counts. Once these mechanistic explanations are available, these oscillations may contribute to real-world outbreak control.

Our study has several limitations. First, the overall epidemic observation period we studied was only 2.5 years, however, the spatiotemporal granularity of the case rate data—by county and by day—is unprecedented, which allowed detailed decomposition with SVD. In addition, the patterns we observed are consistent across multiple subsets of counties in separate states and separate longitudinal bands. Second, case rates are a function of testing rates and test positive rates which could fluctuate depending on the properties of the dominant strains, public awareness and concern, and local testing policies. To confirm the robustness of the patterns derived using case

rates, we showed that these patterns could be closely replicated by analyses of mortality data. Third, we have no way to prove or even expect that the spatiotemporal wave structure we document here in the pandemic will persist into the future. Indeed, the patterns appear to have been disrupted, albeit temporarily, by emergence of the Omicron variant. Nonetheless, further studies of forces that have driven these early-epidemic oscillations may prove important in understanding the fundamental transmission dynamics of COVID-19, as well as other emerging infectious disease epidemics. Fourth, we used the New York Times as the only source of data for this analysis. Confirmation of these patterns using other sources could be done, such as Johns Hopkins University's data, USAFacts, and the COVID Tracking Project. Fifth, because the SVD analysis isolates orthogonal modes from the primary signal, it cannot detect interactions between the EUCO and NUCO oscillators if any exist. Future analyses may reveal interactions between these two modes.

A full understanding of COVID-19 epidemic waves may be especially important given the fact that immunity to COVID-19 wanes[29,30]. Rapid waning of human immunity to SARS-Cov-2, both after infection or vaccination, has led to a public health strategy of frequent boosters[31]. For influenza, annual vaccinations are recommended to be given not earlier than October, to avoid too great a decline in immunity before the expected winter influenza surge[32]. Similarly, SARS-CoV-2 vaccinations mightideally be timed to just precede the expected epidemic surge for any given US region.

Methods

Data sources

Daily case rates were obtained from The New York Times. For all analyses, we only included the 48 continental US states from the beginning of the epidemic in the USA through August 15, 2022. The total observation period is 937 days because the first reported case in the data is January 21, 2020. For county analyses we included only those counties within the 48 continental US states (total N = 3108). The total county data set included $937 \times 3108 = 2.9$ million data points, which aggregated 150 million continental USA reported cases by day and county. In all the analyses, we used log case rates per 100,000 people since the case rates follow a log-normal distribution. To smooth the trends, we used 14 day moving average of case rates before taking the log. We chose 14 days instead of 7, because during the latter part of the observation period, some counties reported cases on a bi-weekly basis. In addition, before taking the log of case rates, we added one to the case rate per 100,000 to avoid numerical issues with undefined log of zero.

Analyses

State level analyses

We conducted geo-clustering analysis of states based on their COVID-19 epidemic pattern similarities, by comparing state vs state daily case rates time series over 937 days. First, we used a simple approach of hierarchical clustering analyses of state-level log case rates. For the distance metric, we used $1 - r$, where r is the Pearson correlation coefficient in log case rates over the entire period between each pair of states. The Pearson correlation coefficient r is calculated as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

where x_i and y_i are the log case rates for states x and y on day i, and $\overline{x}$ and $\overline{x}$ are the mean log case rates for states x and y , respectively, over the 937 days.

The distance $1 - r$ essentially quantifies the dissimilarity between the states' epidemic patterns: a value of 0 indicates perfect correlation (i.e., identical patterns), while a value of 2 indicates perfect anti-correlation.

For the agglomeration algorithm, we chose Ward's method because it minimizes the total within-cluster variances[33]. Specifically, Ward's method seeks to minimize the increase in the total within-cluster variance, or error sum of squares (ESS), at each step of clustering. The increase in ESS when merging two clusters A and B into a new cluster C is given by:

$$\Delta ESS = \sum_{i \in C}(x_i - \overline{x}_C)^2 - \left(\sum_{i \in A}(x_i - \overline{x}_A)^2 + \sum_{i \in B}(x_i - \overline{x}_B)^2\right)$$

where x_i are the data points, and $\overline{x}_A$, $\overline{x}_B$, and $\overline{x}_C$ are the centroids of clusters A, B, and C, respectively. We selected the top two clusters from the Ward method to compute the ratio of each cluster's average case rate to the average of case rates across all states (see below).

Cross-date state ranking correlation matrix

This analysis reveals a matrix of state rankings between each pair of dates.

To further investigate the change in state ranking over time, we show this matrix as a heatmap. Each pixel in this heatmap represents the state rank correlations (Spearman's correlation coefficient) of case rates across states between each pair of dates indicated on the x and y axes. When the state rankings are similar between two dates, the correlation will be red and when the state rankings are reversed, the correlation will be indicated in blue. This heatmap captures secondary fluctuations/oscillations in the states' relative rankings revealing a "checkerboard" pattern because it reveals periods of times where state rankings were similar vs where rankings were reversed regardless of the time trend or case rate in these two dates.

Definition of COVID epidemic waves. We defined the following epidemic waves based on the intersection of the average case trend ratios from the clustering analyses.

Epidemic wave	Start date	End date	Number of days
1	Mar 20, 2020	Jun 1, 2020	73
2	Jun 1, 2020	Oct 1, 2020	122
3a	Oct 1, 2020	Dec 10, 2020	70
3b	Dec 10, 2020	Mar 5, 2021	85
4	Mar 5, 2021	Jun 1, 2021	88
5	Jun 1, 2021	Sep 20, 2021	111
6a	Sep 20, 2021	Jan 1, 2022	103
6b	Jan 1, 2022	Feb 20, 2022	50
7	Feb 20, 2022	Jun 10, 2022	110
8	Jun 10, 2022	Aug 15, 2022	66

Average number of days across waves = 87.8 days.

County-level analyses

Singular value decomposition (SVD) is a widely used technique in linear algebra to obtain low-rank orthogonal modes from data[34]. We used SVD to decompose county-level log case rates data. Mathematically, SVD decomposes a matrix $\boldsymbol{X}$ into three matrices:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\boldsymbol{T}},$$

where: X is an $m \times n$ matrix of county-level log case rates, with x_{jt} representing the log case rate for county j at time t. U is an $m \times m$ orthogonal matrix representing the spatial weights, also known as eigenmaps. Each column k of U corresponds to the left singular vector for the kth mode. Σ is an $m \times n$ diagonal matrix of singular values, where σ_k represents the singular value for the kth mode. The singular values quantify the influence of each mode. V^T is an $n \times n$ orthogonal matrix representing the temporal trends, also known as eigen-trends. Each row k of V^T corresponds to the right singular vector for the kth mode.

The SVD decomposes the spatiotemporal patterns of case rates into a set of modes that, when summed together, reproduce the log case trends. The influence of each mode is computed as the proportion of the total variance explained by each mode k:

$$\text{Proportion of Variance Explained} = \frac{\sigma_k^2}{\sum_i \sigma_i^2}$$

where σ_k^2 is the square of the k-th singular value, and $\sum_i \sigma_i^2$ is the sum of the squares of all singular values. This proportion indicates how much of the total variance in the data is captured by each mode, allowing us to identify the most significant patterns in the county-level COVID-19 case rates.

Data availability

COVID-19 cases and death data are publicly available from the New York Times and is available on GitHub at https://github.com/nytimes/covid-19-data.

Received: 1 May 2024; Accepted: 9 September 2024
Published online: 16 September 2024

References

1. *Coronavirus (Covid-19) Data in the United States*. (The New York Times, 2021).
2. CDC. *COVID Data Tracker Weekly Preview*, https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html (2021).
3. Castro, M. C. *et al.* Spatiotemporal pattern of COVID-19 spread in Brazil. *Science* **372**, 821–826. https://doi.org/10.1126/science.abh1558 (2021).
4. Kim, S., Kim, M., Lee, S. & Lee, Y. J. Discovering spatiotemporal patterns of COVID-19 pandemic in South Korea. *Sci. Rep.* **11**, 24470. https://doi.org/10.1038/s41598-021-03487-2 (2021).
5. Nazia, N., Law, J. & Butt, Z. A. Identifying spatiotemporal patterns of COVID-19 transmissions and the drivers of the patterns in Toronto: A Bayesian hierarchical spatiotemporal modelling. *Sci. Rep.* **12**, 9369. https://doi.org/10.1038/s41598-022-13403-x (2022).
6. Zhu, D., Ye, X. & Manson, S. Revealing the spatial shifting pattern of COVID-19 pandemic in the United States. *Sci. Rep.* **11**, 8396. https://doi.org/10.1038/s41598-021-87902-8 (2021).
7. Oswin, H. P. *et al.* The dynamics of SARS-CoV-2 infectivity with changes in aerosol microenvironment. *Proc Natl Acad Sci U S A* **119**, e2200109119. https://doi.org/10.1073/pnas.2200109119 (2022).
8. Kilgour, E., Rankin, N., Ryan, S. & Pack, R. Mucociliary function deteriorates in the clinical range of inspired air temperature and humidity. *Intensive Care Med* **30**, 1491–1494. https://doi.org/10.1007/s00134-004-2235-3 (2004).
9. Pitzer, V. E. *et al.* Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics. *Science* **325**, 290–294. https://doi.org/10.1126/science.1172330 (2009).
10. Cummings, D. A. *et al.* Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature* **427**, 344–347. https://doi.org/10.1038/nature02225 (2004).
11. Garcia-Calavaro, C. *et al.* North to south gradient and local waves of influenza in Chile. *Sci Rep* **12**, 2409. https://doi.org/10.1038/s41598-022-06318-0 (2022).
12. Bjørnstad, O. N. & Grenfell, B. T. Noisy clockwork: time series analysis of population fluctuations in animals. *Science* **293**, 638–643. https://doi.org/10.1126/science.1062226 (2001).
13. Dutta, A. COVID-19 waves: variant dynamics and control. *Sci Rep* **12** (2022). https://doi.org/10.1038/s41598-022-13371-2

14. Dolgin, E. vaccine immunity is waning-how much does that matter?. *Nature* **597**, 606–607. https://doi.org/10.1038/d41586-021-02532-4 (2021).
15. Catching, A., Capponi, S., Yeh, M. T., Bianco, S. & Andino, R. Examining the interplay between face mask usage, asymptomatic transmission, and social distancing on the spread of COVID-19. *Sci Rep* **11**, 15998. https://doi.org/10.1038/s41598-021-94960-5 (2021).
16. Coccia, M. Meta-analysis to explain unknown causes of the origins of SARS-COV-2. *Environ Res* **211**, 113062. https://doi.org/10.1016/j.envres.2022.113062 (2022).
17. Merow, C. & Urban, M. C. Seasonality and uncertainty in global COVID-19 growth rates. *Proc. Natl. Acad. Sci. USA* **117**, 27456–27464. https://doi.org/10.1073/pnas.2008590117 (2020).
18. D'Amico, F. *et al.* COVID-19 seasonality in temperate countries. *Environ. Res.* **206**, 112614. https://doi.org/10.1016/j.envres.2021.112614 (2022).
19. Smit, A. J. *et al.* Winter is coming: A southern hemisphere perspective of the environmental drivers of SARS-CoV-2 and the potential seasonality of COVID-19. *Int. J. Environ. Res. Public Health* **17** (2020). https://doi.org/10.3390/ijerph17165634
20. Yin, C., Zhao, W. & Pereira, P. Meteorological factors' effects on COVID-19 show seasonality and spatiality in Brazil. *Environ. Res.* **208**, 112690. https://doi.org/10.1016/j.envres.2022.112690 (2022).
21. Zoran, M. A. *et al.* Assessing the impact of air pollution and climate seasonality on COVID-19 multiwaves in Madrid, Spain. *Environ. Res.* **203**, 111849. https://doi.org/10.1016/j.envres.2021.111849 (2022).
22. Hoogeveen, M. J., Kroes, A. C. M. & Hoogeveen, E. K. Environmental factors and mobility predict COVID-19 seasonality in the Netherlands. *Environ. Res.* **211**, 113030. https://doi.org/10.1016/j.envres.2022.113030 (2022).
23. Rendana, M. Impact of the wind conditions on COVID-19 pandemic: A new insight for direction of the spread of the virus. *Urban Clim.* **34**, 100680. https://doi.org/10.1016/j.uclim.2020.100680 (2020).
24. Mercola, J., Grant, W. B. & Wagner, C. L. Evidence regarding vitamin D and risk of COVID-19 and its severity. *Nutrients* https://doi.org/10.3390/nu12113361 (2020).
25. Reich, N. G. *et al.* Collaborative hubs: Making the most of predictive epidemic modeling. *Am. J. Public Health* **112**, 839–842. https://doi.org/10.2105/AJPH.2022.306831 (2022).
26. Rosenfeld, R. & Tibshirani, R. J. Epidemic tracking and forecasting: Lessons learned from a tumultuous year. *Proc. Natl. Acad. Sci. USA* **118** (2021). https://doi.org/10.1073/pnas.2111456118
27. Moriyama, M., Hugentobler, W. J. & Iwasaki, A. Seasonality of respiratory viral infections. *Annu. Rev. Virol.* **7**, 83–101. https://doi.org/10.1146/annurev-virology-012420-022445 (2020).
28. Borchers, A. T., Chang, C., Gershwin, M. E. & Gershwin, L. J. Respiratory syncytial virus–a comprehensive review. *Clin. Rev. Allergy Immunol.* **45**, 331–379. https://doi.org/10.1007/s12016-013-8368-9 (2013).
29. Tregoning, J. S., Flight, K. E., Higham, S. L., Wang, Z. & Pierce, B. F. Progress of the COVID-19 vaccine effort: Viruses, vaccines and variants versus efficacy, effectiveness and escape. *Nat. Rev. Immunol.* **21**, 626–636. https://doi.org/10.1038/s41577-021-00592-1 (2021).
30. Cevik, M., Grubaugh, N. D., Iwasaki, A. & Openshaw, P. COVID-19 vaccines: Keeping pace with SARS-CoV-2 variants. *Cell* **184**, 5077–5081. https://doi.org/10.1016/j.cell.2021.09.010 (2021).
31. Burki, T. Booster shots for COVID-19-the debate continues. *Lancet Infect. Dis.* **21**, 1359–1360. https://doi.org/10.1016/s1473-3099(21)00574-0 (2021).
32. Krammer, F. *et al.* Influenza. *Nat. Rev. Dis. Primers* **4**, 3. https://doi.org/10.1038/s41572-018-0002-y (2018).
33. Ward, J. H. Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
34. Peters, T. *Data-driven science and engineering: Machine learning, dynamical systems, and control: by SL Brunton and JN Kutz, 2019, Cambridge, Cambridge University Press, 472 pp.,£ 49.99 (hardback), ISBN 9781108422093. Level: Postgraduate. Scope: Textbook*. Vol. 60 (Taylor & Francis, 2019).

Acknowledgements

We thank Ben Kirtman, Professor of Atmospheric Sciences and Director of the Cooperative Institute for Marine & Atmospheric Studies, University of Miami, for suggesting use of the SVD method of analysis.

Author contributions

Conceptualization: HJ, DSB Methodology: HJ Investigation: HJ, KL Visualization: HJ Funding acquisition: HJ, DSB Project administration: HJ Supervision: HJ, DSB Writing-original draft: HJ, DSB Writing-review & editing: HJ, DSB, KL.

Funding

HJ was supported by Canada Research Chairs (Award No. CRC-2021-00354).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.